

Removing Weather-Related Image Degradation at the Edge

Ramit Pahwa Computer Vision @ Rivian



Motivation



- Captures from vehicles are degraded due to adverse weather, this adversely affects computer vision algorithms like detection, segmentation and depth estimation.
- A clean image is also desirable in photography.









Mathematical Representation



Raindrop

$$I=(1-M)\odot B+R$$

- I = Degraded Image
- B= Background Image
- R = Residual Image
- M = Mask of Raindrop





Mathematical Representation



Rain Streaks

$$I = \left(B + \sum_{i}^{n} R_{i}\right) \odot T + (1 - T) \odot A$$

- I = Degraded Image
- B= Background Image
- R = Residual Image







Mathematical Representation



Snow

- $I = (M) \odot S + (M) \odot (1-z)$
- S = Snowflakes
- z = Snow Mask
- M = Mask of Snow



Snowy Image



Snow Mask



Metrics: PSNR and SSIM



PSNR

$$MSE = rac{1}{m\,n}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j)-K(i,j)]^2.$$

The PSNR (in dB) is defined as

j,

$$egin{aligned} PSNR &= 10 \cdot \log_{10} \left(rac{MAX_I^2}{MSE}
ight) \ &= 20 \cdot \log_{10} \left(rac{MAX_I}{\sqrt{MSE}}
ight) \ &= 20 \cdot \log_{10} (MAX_I) - 10 \cdot \log_{10} (MSE). \end{aligned}$$

SSIM

$$ext{SSIM}(x,y) = rac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)}$$

with:

- μ_x the pixel sample mean of x;
- μ_y the pixel sample mean of y;
- σ_x^2 the variance of x;
- σ_y^2 the variance of y;
- σ_{xy} the covariance of x and y;

• $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with weak denominator;

- L the dynamic range of the pixel-values (typically this is $2^{\#bits \; per \; pixel} 1$);
- $ullet k_1=0.01$ and $k_2=0.03$ by default.

🚸 RIVIA N

Metrics NR-IQA



NIQE[1-5]



MUSIQ[0-100]



ChatGPT Argues About Weather?





What is image degradation present in the image and its severity on a scale of 1-100?

ChatGPT

The image appears to have water droplets as an overlay, simulating the effect of looking through a wet glass surface. The effect does not completely obscure the scene, but it does add a layer of distortion. If we were to rate the degradation due to these water droplets on a scale from 1 to 100, with 1 being almost imperceptible and 100 being completely degraded, this might be around a 30. The scene behind the droplets remains largely identifiable, but the droplets significantly alter the texture and clarity of the image.

<>> ℃ ℃ ♡ ** -



Multi Modal LLM

Explain and rate degradation



Image Enhancement Timeline





📀 RIVIA N

Image Enhancement: Encoder Decoder



- We can use Encoder-Decoder U-net Network [1] to filter out distortions.
- Various methods are developed which train this U-net on specific types of degradation raindrop, haze etc. Each will have a specific model [2][3].





Finetuning Specific Task



- E1 D1 E2 D2 D3 **E3**
- Most of these methods just focus on one task at hand or fine-tune the model separately for each task.
- Adopting this into real time systems is complex.

Next step: Can we have the same network restore all degradation together?



Image Enhancement : All in One



- All-in-one weather [4] restoration approach uses a single network for all restoration types.
- We can learn a specific encoder for different weather degradation using neural architecture search.
- Generic decoder which is shared across different encoders.





Image Enhancement : TransWeather[5]



- Transformer network based on DETR[6] hybrid architecture.
- Learns a single encoder + single decoder with learned weather queries.
- Weather queries? So, these enable different neurons to activate for different degradation.
- Latency of the transformer model is several magnitudes greater than what can be supported on edge (100 GB, 200 ms/image on T4). Solution?



🚸 RIVIA N

Choice of Encoder/ Decoder

1





We found transformer encoder performed better as it can capture global context. We extract intermediate features and stack them before passing to the decoder.

- Deconvolution Decoder 3.

Convolution decoder coupled with upscaling layer gives us the best FID without checkerboard artifacts. Transformers perform similar with added latency.

Hybrid architectures are the way to go for de-weathering; choice depends on the task



Image Enhancement : TransWeather[5]





15

Image Enhancement Models : Training

- Training of the Encoder Decoder architecture is done using two losses:
- Reconstruction Loss: L1, L1-smooth, L2, etc.
- Perceptual Loss: Extract features of the output and ground truth using VGG.

$$\mathcal{L}_{smoothL_1} = \begin{cases} 0.5\mathbf{E}^2 & \text{if } |\mathbf{E}| < 1\\ |\mathbf{E}| - 0.5 & \text{otherwise} \end{cases},$$
$$\mathbf{E} = \hat{\mathbf{I}} - \mathbf{G}$$

 $\mathcal{L}_{perceptual} = \mathcal{L}_{MSE}(VGG_{3,8,15}(\hat{\mathbf{I}}), VGG_{3,8,15}(\mathbf{G})).$





Image Enhancement Models : Training

- Training of the Encoder Decoder architecture is done using two losses:
- Reconstruction Loss: L1, L1-smooth, L2 etc.
- Perceptual Loss: Extract features of the Output and ground truth using VGG.

RIVIAN

$$\mathcal{L}_{smoothL_1} = \begin{cases} 0.5\mathbf{E}^2 & \text{if } |\mathbf{E}| < 1\\ |\mathbf{E}| - 0.5 & \text{otherwise} \end{cases},$$
$$\mathbf{E} = \hat{\mathbf{I}} - \mathbf{G}$$

 $\mathcal{L}_{perceptual} = \mathcal{L}_{MSE}(VGG_{3,8,15}(\hat{\mathbf{I}}), VGG_{3,8,15}(\mathbf{G})).$

$$\mathcal{L}_{total} = \mathcal{L}_{smoothL_1} + \lambda \mathcal{L}_{perceptual},$$



Language Aided Weather Restoration



- Combining language prior with mixture of experts model for best performance.
- Introduction of degradation prior, enhancing model's understanding of weather type and severity from images.
- Degradation prior used to guide the sparse selection of experts, ensuring adaptive response to varied conditions



Figure 1. Score (PSNR and SSIM) comparisons. We compare our model (red) with the top 2 (blue and green) baselines on benchmark datasets with various weather scenarios. Superscripts besides evaluation metrics are used to differentiate benchmark datasets.



Language Driven Model[8]









- There are multiple open-source datasets available. Raindrop[7][8], snow [9], rain streak[10][11].
- Size of the datasets is very small; the biggest dataset for realistic raindrop for example is only ~6000 annotated clean, degraded pairs.
 - Hard to capture degraded sequence and corresponding clean sequence.
 - Carefully capture with specialized hardware.



• The generalizability of the networks trained on open-source dataset doesn't translate well on Automotive Fish-eye captures



Datasets: Realistic Synthetic Creation



- Ground truth data is hard to come by, so we trained cycle GANs to generate realistic weather degraded sequence.
- We first collect an open-source dataset and also synthetically create weather degradation of different types.





Datasets: Realistic Synthetic Creation



- **Unsupervised** training of Haze vs Non-Haze image
- We obtain two networks: G, which can dehaze and F, which can produce haze.
- Empirically the dehazing model is low quality and often suffers from hallucinations.
- We can use F to create weather degraded sequences for training



Cvcle-Consistency + Perceptual-Consistency

Supervised Restoration is Better



- Supervised training of GAN's inspired from [2] works well on when trained on the dataset we created.
- Compress the network to reasonable size to be able to run on edge [Qualcomm 8155 NPU acceleration.]
 - 1. GAN compression[12] gives us an easy process for compressing our trained GAN's.
 - 2. Knowledge distillation from original teacher network to smaller student network.
 - 3. Computation: 9-21x less
 - 4. Memory: 4-33x better

RIVIAN



Accelerating Horse2zebra by GAN Compression



Original CycleGAN; FLOPs: 56.8G; FPS: 12.1; FID: 61.5



GAN Compression; FLOPs: 3.50G (16.2x); FPS: 40.0 (3.3x); FID: 53.6

(c) Rivian

Results



Input



Input



Output



Output



GT



GT





Improved Object Detection







Deploying on the Edge : SNPE









Runtime Format: ONNX



 ONNX is an open ML standard which allows model to run across gamut of hardware, from cloud to edge devices.

```
torch_model = MyModel()
torch_input = torch.randn(1, 1, 32, 32)
onnx_program = torch.onnx.dynamo_export(torch_model, torch_input)
```





Runtime Format : .dlc



- .dlc has some supported layers and operations
 - Network layers: Link
 - ONNX ops : <u>Link</u>
- Caveat: Many newer models have custom operations and often cannot be converted to .dlc file.
- snpe-pytorch-to-dlc --input_network resnet18.pt --input_dim input "1,3,224,224" --output_path resnet18.dlc
- snpe-onnx-to-dlc --input_network models/bvlc_alexnet/bvlc_alexnet/model.onnx --output_path bvlc_alexnet.dlc



Conclusion



- Models are data hungry; we produced a scalable way to create realistic degradation.
- Use simple encoder-decoder architecture to restore images.
- Leverage SNPE to run inference on edge devices.
- Language driven approaches although provide the best results, adapting them on edge is complex.



Resources



- Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. "U-net: Convolutional networks for biomedical image segmentation." *Medical image computing and computer-assisted intervention– MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18.* Springer International Publishing, 2015.
- 2. Qian, Rui, et al. "Attentive generative adversarial network for raindrop removal from a single image." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018.
- 3. Chen, Sixiang, et al. "SnowFormer: Context interaction transformer with scale-awareness for single image desnowing." *arXiv preprint arXiv:2208.09703* (2022).
- 4. Li, Ruoteng, Robby T. Tan, and Loong-Fah Cheong. "All in one bad weather removal using architectural search." *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 2020.



Resources



- 5. Valanarasu, Jeya Maria Jose, Rajeev Yasarla, and Vishal M. Patel. "Transweather: Transformer-based restoration of images degraded by adverse weather conditions." *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2022.
- 6. Carion, Nicolas, et al. "End-to-end object detection with transformers." *European conference on computer vision*. Cham: Springer International Publishing, 2020.
- Stefan Roth and Michael J Black. Fields of experts: A framework for learning image priors. In 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), volume 2, pages 860–867. IEEE, 2005.
- 8. Yang, Hao, et al. "Language-driven All-in-one Adverse Weather Removal." *arXiv preprint arXiv:2312.01381* (2023).
- 9. SNPE Resources: Tutorial on conversion : <u>Link</u>, Visualizer : <u>Link</u>, Details on quantization : <u>Link</u>

